Flow Direction Toolkit: Matlab Scripts for Flow Direction Corrections

Jacob Schaperow

June 12, 2019

**Abstract**

Hydrologic routing models require flow direction data to route spatially-distributed runoff through river channels to the basin outlet. However, the spatial resolution of the routing model is generally much coarser than the DEM used to generate the flow direction map. Upscaling the DEM or the flow direction map introduces error in the river network and basin area. Thus, hydrologic modelers have to make manual corrections to the flow direction map before using the routing model. The Flow Direction Toolkit is a Matlab toolbox designed to make manual correction of flow direction maps relatively pain-free. It has two functions: it can be used to make manual flow direction corrections, or it can incorporate knowledge of the true river network to automatically adjust the flow direction map. We demonstrate the capability of the Flow Direction Toolkit by adjusting a coarse-resolution flow direction map to better match the basin area and river network from a fine resolution DEM.

**Introduction**

Routing models (e.g. Lohmann et al., 1998) require flow direction data for routing runoff from each grid cell in the river basin to the outlet. Flow directions define the river network. Usually, flow directions are calculated automatically from a digital elevation model (DEM) based on the elevation gradient between adjacent cells. The two most commonly used algorithms are eight-directional (D8) and multidirectional (D-Inf). Most routing models require a flow direction map generated using a D8 algorithm.

The accuracy of flow direction maps based on DEMs depends on the resolution and vertical accuracy of the DEM. Coarse resolution means that narrow river channels may not be resolved, or there may be more than one river channel in one grid cell. Vertical accuracy of the DEM comes into play for flat areas where topographic relief is small. Flow direction maps also struggle in their representation of lakes.

DEMs are available globally at quite high resolution. One example is the SRTM-based MERIT DEM (Yamazaki et al., 2017), available at 3 arc-second, or approximately 90 m resolution. Flow direction maps derived from high resolution DEMs have less error than those derived from coarser resolution DEMs (for instance Hydro1k). However, often, hydrologists want to run routing models at coarse resolution because of the computation cost and increased input data requirements to run a distributed hydrological model at very fine resolution. Thus, there is a need for coarse resolution flow direction maps that have drainage areas and river networks consistent with the more accurate, fine resolution DEMs.

Various researchers have tried to tackle this spatial upscaling problem (Wu et al., 2012; Yamazaki et al., 2009), but in the end, there is still always a need for some manual corrections. This software is designed to help the user perform these manual corrections to get a better match between the coarse-resolution flow direction map and the true river network. The true river network is assumed known. For example, if a fine resolution DEM is available, then the river network derived at fine resolution could be used as the "true" river network.

The Flow Direction Toolkit is a set of Matlab functions for assessing the quality of automatically-generated flow direction maps and making manual corrections as necessary to get good agreement between flow direction maps and known basin boundaries/river channel locations. Given a flow direction map and a known basin boundary and river network, the Flow Direction Toolkit makes plots showing the basin, river networks, and flow directions and allows the user to click on the map to find the coordinates of the cell whose flow direction needs to be changed. We suggest making a list of coordinates to change in an Excel spreadsheet, and making all the changes at once. Then, the modified flow direction map can be plotted alongside the river network to double-check that the corrections are appropriate. In this document, we descibe the methods used and show several case studies where we perform manual flow direction corrections for differently-sized river basins.

**Methods**

To illustrate our methods, we use a made-up test case (Figure 1). In practice, the flow direction map would be derived from a DEM. The flow direction map does not agree with the known river network (Figure 1, in red). In addition to plotting the basin boundary, flow directions, river network, and grid cell outlines, the Flow Direction Toolkit checks each grid cell for three conditions:

1. Does the cell flow off the edge of the domain?
2. Does the flow direction agree with the direction of the known river?
3. Is the cell part of an infinite loop where two cells flow into one another?

If any of these conditions are true, the Flow Direction Toolkit flags the cell on a map so the user can focus on these cells for manual corrections. If the flow direction map is truly in agreement with the basin boundary, then the flow direction should only point outside the basin boundary in one cell: the outlet.

*Flow direction of river cells*

Although it is not possible to retrieve the flow direction map used to generate the river network because the river network is the result of aggregating flow direction data together, if there is a known river network, then we can back out some information about the flow direction in the cells through which the river flows (river cells). We used the following method to determine the flow direction of the river cells.

The river data should be stored as a shapefile, with the river vectors listed from upstream to downstream. The algorithm assumes that the length of the river vectors should increase from upstream to downstream and will automatically sort the river vector by length. This assumption is not generally true, but it is easy to implement and we think that segment length should correlate with downstream distance because upstream tributaries tend to be relatively short compared to rivers downstream. In the future, this should be replaced with an algorithm that actually does sort the river segments from upstream to downstream, such as the calculate\_chainage() Python function from Kostas Andreadis.

The goal is to assign an angle theta to each river cell. Given a gridded study area and a river network shapefile, there are four possibilities:

1. The river channel does not pass through this grid cell
2. The river channel passes through the cell, but there are no vertices located in the cell
3. The river channel passes through the cell, but there is only one vertex in the cell
4. The river channel passes through the cell, and there are two or more vertices in the cell

The Flow Direction Toolkit handles each of these possibilities. If the river channel does not pass through the grid cell, then no flow direction can be inferred. If the river channel passes through the cell, but there are no vertices or only one vertex located in the cell, then the portion of the river segment is subdivided into several smaller river segments in order to guarantee that there are at least two vertices in the cell. Given that there are two or more vertices in the cell, then the first and last vertices are used to calculate the flow direction, assuming the flow starts at the first vertex and flows to the last vertex.

Again, the flow direction might be backwards if the river vectors are list from downstream to upstream instead of upstream to downstream, and the derived flow directions are predisposed to be diagonal, even when they should not be, just because of the way theta is calculated. For these reasons, we do not recommend using the automatic flow direction corrections, and the flagged "disagrees with river data" cells should be taken with several grains of salt.

*Infinite flow direction loops*

Manual modifications to the flow direction map can introduce infinite loops, where flow is routed in circles forever and never makes it to the basin outlet. The Flow Direction Toolkit highlights any two adjacent cells whose flow goes in an infinite loop. These so-called "trivial loops" are easy to spot. Nontrivial loops involving more than two grid cells are harder to identify. Mu Xiao has written a script to identify nontrivial loops. This feature should be included in a later version of the toolkit.

**Example - Upper Tuolumne Basin/automatic corrections**

We ran the Variable Infiltration Capacity (VIC; Liang et al., 1994) model over the Upper Tuolumne basin for the calendar years 2006-2011 in water balance mode at 1/16 degree resolution with a 3-hour timestep. VIC is a land surface model that uses soil and vegetation data to determine how precipitation partitions into runoff, infiltration, evaporation, and storage. VIC produces fields of runoff and baseflow at each grid cell in the model domain. The UW routing model (Lohmann et al., 1998) routes these generated runoff and baseflow fields through the channel network, producing river discharge estimates at specified locations along the river channel. The flow direction map is a critical input for the routing model. Here, we show why the automatically-generated flow directions map from HydroSHEDS is inadequate for 1/16 degree resolution modeling, and that by modifying the flow directions using the Flow Direction Toolkit, we can run the routing model to estimate discharge more accurately.

We demonstrate a workflow using the Flow Direction Toolkit by preparing a coarse resolution flow direction map for portion of the Upper Tuolumne (UT) river basin, located in the Sierra Nevada range in California. The outlet is at O'Shaughnessy Dam, which dams the Hetch Hetchy reservoir. We obtained a DEM at 90 m resolution from hydrologically-conditioned, void-filled SRTM data from the HydroSHEDS hydrography database (Lehner et al., 2008), resampled it to 1 km resolution, filled sinks, and calculate flow directions using the r.watershed function in GRASS GIS (Figure 2). The basin boundary and river centerlines (red) come from HydroSHEDS and are taken as truth data.

Table 1. Inputs to Flow Direction Toolkit.

|  |  |
| --- | --- |
| Coarse-resolution flow directions | Delineated from 5 km HydroSHEDS DEM |
| Basin boundary | Delineated from 1 km HydroSHEDS DEM |
| River centerlines | Extracted from 1 km HydroSHEDS DEM |
| Flow directions | Calculated from 1 km HydroSHEDS DEM |
| Gage locations | Hetch-Hetchy reservoir outlet |

Figure 3 shows a zoomed-in view of the UT basin, showing the flow directions before and after automatically adjusting the flow directions based on the flow direction of the river cells. After the correction, the flow direction map is in better agreement with the HydroSHEDS river centerlines.

**Delineating the Indus River Basin**

MERIT DEM (Yamazaki et al., 2017) is an error-corrected version of the SRTM DEM. We obtained the MERIT DEM (Yamazaki et al., 2017) at its original resolution (3 arc-seconds), and we aggregated it to 30 arc-second and 1/16 degree resolution using GDAL. Then, we used GRASS GIS to delineate the IRB using a couple of different methods.

*Procedure for delineating watershed*

1. Fill sinks using r.hydrodem
2. Calculate flow directions and flow accumulation using r.watershed. Set a threshold of approximately 400 km2 of upstream area to define river channel pixels. (Approximate because I am assuming that 30 arc-seconds is equal to 1000 m).
3. Set outlet coordinates visually by looking for the pixel with the highest flow accumulation value.
4. Delineate the watershed using r.stream.basins
5. Extract the stream network (as a vector file) using r.stream.extract
6. Calculate basin area using v.report

We followed this procedure for both the 30 arc-second and 1/16 degree DEMs, and we obtained somewhat different results (Figure 4). Choosing the outlet coordinate was not a straightforward task. The true outlet location is not the same as the point of highest flow accumulation, due to errors in the DEM (either from measurements or introduced by upscaling). The approximate watershed can still be delineated by "snapping" the outlet location from its true location to a nearby point on the main river channel. Using the point of highest flow accumulation (69.0717° E, 24.0040° N) as the outlet, the 30 arc-second watershed looks like this (Figure 5). The outlet location is not exactly correct, nor are the channel centerlines. The basin area is approximately 1.11 million km2. Since the channel connectivity is incorrect, the upstream area above each dam will not be correct. Dams that are not on the delineated river channel should be snapped to the river channel to ensure that there are on the main flow path.

For the 1/16 degree DEM, the true outlet and the point of highest flow accumulation were different, as well (Figure 6). The point of highest flow accumulation, which was chosen for the outlet, was approximately 68.9693° E, 24.0930° N. The delineated watershed is shown in Figure 7. Again, the stream networks do not perfectly match GRWL. The calculated basin area is 1.13 million km2. Figure 8 shows the difference between the watershed delineated at 30 arc-seconds and the watershed delineated at 1/16 degree resolution.

**Upper Indus Basin**

The basin area for the 3 arc-second MERIT hydrography data is 870,500 km2, considerably smaller than the other estimates in the literature (closer to 1.2 million km2).

Use the 3 arc-second basin as a mask for the 1/16 degree flow directions.

Correction procedure:

1. Make an Excel document called fdir\_corrections.xls
2. Click on Matlab plot to get coordinates and copy them in to the Excel document
3. Type the new flow direction for this cell
4. Repeat for all cells needing flow direction corrections. It is convenient to start from the outlet.
5. Implement changes in correct\_flowdir.m
6. Check that results are as desired using plotflowdir.m
7. Output into a format compatible with GRASS GIS
8. Delineate watershed from modified flow directions as a check

We took the full-resolution MERIT hydrography dataset (3 arc-seconds) and delineated the Indus River Basin using the hydrologically-adjusted elevation data. Figure 1 shows the flow accumulation map, with dams and GRWL centerlines overlaid. Figure 2 shows the delineated basin, with the outlet located at \_\_\_. The basin area is \_\_\_ million km2. This is all well and good, but we need a 1/16 degree resolution flow direction map that matches the true basin shape and flow network. Since we think that the 3 arc-second basin is pretty good, we can use the Flow Direction Toolkit to adjust the 1/16 degree flow direction map to better match the 3 arc-second flow direction map. Another option is to directly upscale the 3 arc-second map using a method like Dominant River Tracing (Wu et al., 2012) or FLOW (Yamazaki et al., 2009). But the results of this would likely need to double checked, as well. It is entirely possible that they are also unsuitable for using as inputs to the routing model. The 1/16 degree DEM compares the best to the IRB map in the FAO Aquastat report (FAO, 2011).

The basin area of the corrected IRB is 817,400 km2. While smaller than other estimates, we believe this to be the correct value. Previous estimates of the basin area relied on less accurate DEMs.

**Discussion**

The MERIT Hydrography dataset (Yamazaki et al., 2019) was designed to minimize the need to manual corrections, but it should still be checked, as well. [Do a case study.]

Comparison with HydroSHEDS, Wu et al, etc.

The algorithm for finding the direction of the river cells is not perfect. It tends to assume that there are too many diagonal cells. Improvements could be made, for example, using techniques from Yamazaki et al. (2009).

Future improvements: nontrivial loops, computational efficiency, etc.

Since the automated flow direction corrections are still very much a work in progress, we demonstrate the manual flow direction process in this section, for the Example Basin.

**References**

Lohmann et al., 1998)

Yamazaki et al. (2019)

Yamazaki et al. (2009)

Wu et al. (2012)

Yamazaki et al., 2017

Other citations from Yamazaki papers

FAO. 2011. http://www.fao.org/nr/water/aquastat/basins/indus/indus-CP\_eng.pdf

Lehner, B., Verdin, K., Jarvis, A. (2008): New global hydrography derived from spaceborne elevation data. Eos, Transactions, AGU, 89(10): 93-94.

**Submission notes**

There are numerous issues with the algorithm that need to be addressed. Need to make it very clear that this is making a contribution to science. Definitely needs more rigorous testing before I can submit it. Ideally, should do a test all the way from basin delineation through running the routing model. Can use Livneh et al. (2013) VIC outputs to run the routing model.

Can test the code for the following cases:

Made-up test basin - shows what elements can be plotted, and how FDT can be used to adjust the flow direction for individual grid cells. Could go all the way with this example (running VIC), but then we would need to set up a realistic "fake" VIC setup.

Upper Tuolumne Basin - shows 30 arc-second flow direction map for the UT basin and how the river network can be used to automatically adjust the flow directions to match the river network. This test case is interesting because we've actually run VIC for this domain, but at 1/16 degree resolution, the flow direction map is essentially the same as the 30 arc-second map. There appears to be no need to adjust the flow direction map.

Upper Indus Basin - shows how 30 arc-second and 1/16 degree basin delineations can look different. It turns out that the coarse resolution delineation is actually more accurate that the fine resolution delineation, in this case. See Khan et al. (2014) for a discussion of the true shape of the UIB. There is no need to adjust the flow direction map to better match the fine resolution map because the coarse resolution map is already quite good. For the sake of demonstrating FDT, it could be used to add successive (spurious) portions of the UIB by adjusting the flow direction at each of the critical points discussed by Khan et al. (2014). We have run VIC for this domain, so the results could potentially be used in a paper.

Potential journal: Environmental Modelling and Software